Chloroplast genome evolution in the Dracunculus clade (Aroideae, Araceae)

Abdullah

2021-01

Abdullah , Henriquez , C L , Mehmood , F , Hayat , A , Sammad , A , Waseem , S , Waheed , M T , Matthews , P J , Croat , T B , Poczai , P & Ahmed , I 2021 , ' Chloroplast genome evolution in the Dracunculus clade (Aroideae, Araceae) ' , Genomics , vol. 113 , no. 1 , pp. 183-192 . https://doi.org/10.1016/j.ygeno.2020.12.016

http://hdl.handle.net/10138/327039
https://doi.org/10.1016/j.ygeno.2020.12.016

Genomics 113 (2021) 183–192

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

Chloroplast genome evolution in the *Dracunculus* clade (Aroideae, Araceae)

Abdullah [a,*], Claudia L. Henriquez [b], Furrukh Mehmood [a], Amir Hayat [c], Abdul Sammad [d], Shahid Waseem [e], Mohammad Tahir Waheed [a], Peter J. Matthews [f], Thomas B. Croat [g], Peter Poczai [h,*,1], Ibrar Ahmed [e,*,1]

[a] *Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan*
[b] *Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA*
[c] *Department of Biochemistry, Faculty of Life and Chemical Sciences, Abdul Wali Khan University, Mardan, Khyber Pakhtunkhwa, Pakistan*
[d] *Department of Botany, Ghazi University, Dera Ghazi Khan, Pakistan*
[e] *Alpha Genomics Private Limited, Islamabad 45710, Pakistan*
[f] *Department of Cross-field Research, National Museum of Ethnology, Osaka, Japan*
[g] *Missouri Botanical Garden, 4344 Shaw Blvd., St. Louis, MO 63110, USA*
[h] *Finnish Museum of Natural History, University of Helsinki, P.O. Box 7, FI-00014, Finland*

ARTICLE INFO

Keywords:
Dracunculus
Chloroplast genome
Araceae
Phylogeny
Evolution

ABSTRACT

Chloroplast (cp) genomes are considered important for the study of lineage-specific molecular evolution, population genetics, and phylogenetics. Our aim here was to elucidate the molecular evolution in cp genomes of species in the *Dracunculus* clade (Aroideae, Araceae). We report *de novo* assembled cp genomes for eight species from eight genera and also retrieved cp genomes of four species from the National Center for Biotechnology Information (NCBI). The cp genomes varied in size from 162,424 bp to 176,835 bp. Large Single Copy (LSC) region ranged in size from 87,141 bp to 95,475 bp; Small Single Copy (SSC) from 14,338 bp to 23,981 bp; and Inverted Repeats (IRa and IRb) from 25,131 bp to 32,708 bp. The expansion in inverted repeats led to duplication of *ycf1* genes in four species. The genera showed high similarity in gene content and yielded 113 unique genes (79 protein-coding, 4 rRNA, and 30 tRNA genes). Codon usage, amino acid frequency, RNA editing sites, microsatellites repeats, transition and transversion substitutions, and synonymous and non-synonymous substitutions were also similar across the clade. A previous study reported deletion of *ycf1*, *accD*, *psbE*, *trnL-CAA,* and *trnG-GCC* genes in four *Amorphophallus species*. Our study supports conservative structure of cp genomes in the *Dracunculus clade* including *Amorphophallus* species and does not support gene deletion mentioned above. We also report suitable polymorphic loci based on comparative analyses of *Dracunculus* clade species, which could be useful for phylogenetic inference. Overall, the current study broad our knowledge about the molecular evolution of chloroplast genome in aroids.

1. Introduction

The chloroplast (cp) organelle plays an important role in the biosynthesis of carbohydrates, proteins, and lipids in plants [1]. It is composed of a double-stranded genome that shows independent replication from the nuclear genome [1,2]. Chloroplast genomes are mostly quadripartite in structure in which the large-single copy (LSC) region and the small-single copy (SSC) region are separated by a pair of inverted repeats (IRs: IRa and IRb) [2–4]. In some plant lineages, the quadripartite structure is not observed due to loss of one or two IRs [5]. Moreover, very short IRs are also reported in some plant lineages [6]. A mixture of linear and circular cp genomes have also been reported, *e.g.* in maize [7]. The size of cp genomes ranges from 107 kbp (*Cathaya argyrophylla* Chun & Kuang) to 218 kbp (*Pelargonium* × *hortorum* L.H. Bailey) [1]. Many mutational events take place within cp genomes such as small- and large-scale inversions, gene rearrangement, insertions-deletions (InDels), SNPs (single-nucleotide polymorphisms), microstructural variations, and originations of oligonucleotide repeats [1,8–10]. The co-existence among mutational events is also reported in cp genomes [11]. The contraction and expansion of IRs also occur very

* Corresponding authors.
E-mail addresses: abd.ullah@bs.qau.edu.pk (Abdullah), peter.poczai@helsinki.fi (P. Poczai), iaqureshi_qau@yahoo.com (I. Ahmed).
[1] *Authors with equal contribution.*

https://doi.org/10.1016/j.ygeno.2020.12.016
Received 1 July 2020; Received in revised form 19 November 2020; Accepted 10 December 2020
Available online 14 December 2020

commonly in cp genomes, which cause duplication of single-copy genes or convert single copy genes into duplicate genes [12]. Chloroplast genome polymorphisms are used for the study of evolution, population genetics, phylogenetics, and phylogeographics [1,13].

Araceae is a large monocot plant family, comprising 144 accepted genera and 3645 species, although estimations place the number of genera at 150 and species at 6489 [14]. Species of this family include the smallest known angiosperms and occupy circumboreal to tropical habitats [15–17]. The family is divided into eight subfamilies and the subfamily Aroideae is one of the largest subfamilies, comprising 75 genera and 1573 species [15,18]. Among aroids, IR contraction and expansion has either led to duplication of otherwise single-copy genes, or deletion/truncation of one copy of an otherwise duplicated gene [19,20]. Within Aroideae, significant gene rearrangements and IR boundary shifts were found in *Anchomanes hookeri* (Kunth) Schott and *Zantedeschia aethiopica* (L.) Spreng. [19]. Yet, with regards to total gene content, cp genome evolution in aroids has been conservative [19,20]. However, an exception is a recent study reporting deletion of *ycf1*, *accD*, *psbE*, *trnL-CAA*, *trnG-GCC*, *rpl*23 and *rpl*2 genes in four species of genus *Amorphophallus* of the *Dracunculus* clade [21]. Deletion of *rpl*23 and *rpl*2 was limited to only one IR region.

The *Dracunculus* clade is a strongly supported clade within subfamily Aroideae [15,18]. Its members share several anatomical synapomorphies and most have a spinose pollen exine that is otherwise only found in Lemnoideae [15,22]. Ecological, morphological, pollination, and cytological information for the *Dracunculus* clade species included in this study can be found in Table 1 and Fig. 1.

To gain further insight into the evolutionary dynamics of cp genomes in the *Dracunculus* clade, we sampled eight species from three sub-clades (*Amorphophallus*, *Colletogyne*, and *Pistia* clades) (Table 1, Fig. 1), based on phylogenetic inference [15], and report *de novo* assembled cp genomes for these species. We also retrieved four species from NCBI, the accession numbers of whom are given in methodology. With this sample set, our aims were to determine the extent of the gene loss events reported by Liu et al. [21], and to report polymorphic loci that could be useful for further phylogenetic analyses within the *Dracunculus* clade.

2. Materials and methods

2.1. Samples collection, DNA extraction, and sequencing

We selected eight morphologically and ecologically diverse species within the *Dracunculus* clade (Aroideae, Araceae) (Table 1, Fig. 1) based on Cusimano et al. [15]. Fresh and healthy leaf tissues without any apparent disease symptoms were collected for these species from the Araceae Greenhouse at the Missouri Botanical Garden in St. Louis, Missouri (Table 2). The DNA extraction was performed by using the Qiagen DNeasy Minikit (Qiagen, Germantown, MD, USA), with some modifications as described by Henriquez et al. [19]. After confirmation of the quality and quantity of DNA, the libraries were constructed for sequencing in the Pires Lab of the University of Missouri, Columbia, using the protocol of the Illumina TruSeq kit (Illumina, Inc., San Diego, CA, USA). The qualified libraries were sequenced from single-end 100 bp reads using the Illumina HiSeq 2000 at the University of Missouri DNA Core according to Henriquez et al. [19]. The sequencing generated reads from 3.05 GB (11.6 million reads) to 6.41 GB (24.55 million reads) (Table 2).

2.2. Quality of sequencing data, de novo assembly, and annotations

Quality of the short reads was explored using fastQC and MultiQC analyses [23,24], confirming high quality of the data with an average Phred score of 35 and above. (Table 2). All of the genomes were initially assembled by using Velvet v.1.2.0 with various kmer values following Abdullah et al. [8,12]. The assembly by Velvet enabled us to assemble complete cp genomes in 3–8 long contigs. The contigs were imported in

Table 1
Data of chromosomes number, geographical distribution, habitat, morphology, and ecology of eight newly sequenced species.

Species	Inferred ancestor chromosome number	Geographical Distribution	Habit and ecology (Plants of the World Online, powo.science.kew.org, Mayo et al., 1997)
Alocasia fornicata	14	North-east India, North Myanmar, Thailand	medium-sized, clumping, evergreen herbs, stem epigeal erect, with clear to milky late;. found in primary and secondary humid evergreen to semideciduous forests, riverbanks, below 100 m above sea level
Amorphophallus titanum	13	Sumatra, Indonesia	seasonally dormant herbs, often large, sometimes gigantic, tuber usually depressed-globose, found in rainforests of western Sumatra, on steep hillsides, at 120–365 m above sea level
Arisarum simorrhinum	14	South-east Spain, Portugal, Morroco	small, seasonally dormant herbs, stem an ovoid to cylindric tuber with stolons; found in warm temperate scrub and woodland; geophytes, on stony ground in macchie (“maquis”), between rocks or under trees and shrubs
Carlephyton glaucophyllum	27	Madagascar	seasonally dormant herbs, tuber depressed-globose; found in tropical deciduous forest on limestone or basalt; geophytes, in rock crevices and holes with leaf litter, at 150–200 m above sea level
Steudnera colocasiifolia	14	China to Indo-China	medium-sized evergreen herbs, stem epigeal, erect or creeping, covered with fibrous remains of leaves and cataphylls; found in dense forests, wet meadows, by streams
Typhonium blumei	8,13	South China to Indo-China, Nanseishoto to Taiwan	very small to medium-sized evergreen herbs, rhizome broad and flattened; found in warm temperate, subtropical and tropical humid and seasonal forests, secondary forest, cultivated land, among rocks, grasslands
Xanthosoma helleborifolium	13	Colombia, Amazonia, Andes,	evergreen herbs, stem a thick, subcylindric, hypogeal tuber often

(*continued on next page*)

Table 1 (*continued*)

Species	Inferred ancestor chromosome number	Geographical Distribution	Habit and ecology (Plants of the World Online, powo.science.kew.org, Mayo et al., 1997)
		Llanura del Caribe, Pacifico.	producing smaller tubers on stoloniferous side branches; found in tropical moist and humid forest, subtropical forest; geophytes on forest floor, at 30–50 m above sea level
Zomicarpella amazonica	13	Brazil (western Amazonia), Colombia	small evergreen herbs with creeping rhizome found in tropical humid forest ("terra firme"); geophytes on forest floor

*Data of chromosome number has been taken from Cusimano et al. [15] whereas data about geographical distribution, habit, and ecology has been taken from Plants of the World Online, powo.science.kew.org, and Mayo et al. [17].

Geneious Pro (v) to assemble full length cp genomes after manually curating the IR boundary regions, following Abdullah et al. [12]. The assembly of all genomes were further validated by using Fast-Plast pipeline, following Henriquez et al. [25]. Fast-Plast was found to be slightly advantageous over Velvet for some minor corrections at microsatellite loci. Coverage depth analyses were performed by mapping short reads to respective *de novo* assembled genomes through Bowtie [26]. The *de novo* assembled genomes were annotated using GeSeq [27], whereas the tRNA genes were further confirmed by tRNAscan-SE v.2.0 [28] and ARAGORN v.1.2.38 [29]. The circular map of the genomes was drawn by using OrganellarGenomeDRAW (OGDRAW) [30]. These genomes were submitted to GenBank of the National Center for Biotechnology (NCBI) by generating five column tab-delimited tables by GB2Sequin [31] under specific accession numbers (Table 2). The raw data generated in the current study were also submitted to Sequence Read Archive (SRA) under project number PRJNA547613.

2.3. Comparative chloroplast genomics among species of the Dracunculus clade

We compared cp genome features of 12 species from 12 genera (one species from each genus) of the *Dracunculus* clade, including eight species in the current study and four species we previously reported. The previously reported species included *Pinellia pedatisecta* Schott (NCBI accession number MN046890), *Syngonium angustatum* Schott (MN046894), *Arisaema franchetianum* Engl. (MN046885), and *Colocasia esculenta* (L.) Schott (JN105689) [19,32]. Chloroplast genomic features, including gene arrangements, gene content, codon usage, and amino acid frequency, were analyzed in Geneious R8.1 [33]. The RNA editing sites in the cp genomes were analyzed by predictive RNA editor for plant

Fig. 1. Vegetative and/or reproductive stages of included species, and taxonomic ranking within the Dracunculus clade from Cusimano et al. [15]. 1. *Amorphophallus* clade – (A) *Amorphophallus titanum*, Tribe Thomsonieae, (B) *Xanthosoma helleborifolium*, Tribe Caladieae, (C) *Zomicarpella amazonica*, Tribe Caladieae. 2. *Colletogyne* clade – (D) *Arisarum simorrhinum*, Tribe Arisareae, (E) *Carlephyton glaucophyllum*, Tribe Arophyteae. 3. *Pistia* clade – (F) *Steudnera colocasiifolia*, (G) *Alocasia fornicata*, (H) *Typhonium blumei*, Tribe Areae.

Table 2
Quality and quantity of the generated sequences data and GenBank accession number of eight species.

Sample Name	Data in GB	WGS reads (millions)	Average Phred score	Chloroplast reads (millions)	Average coverage	Maximum coverage	GenBank Accession
Alocasia fornicata	6.03	23.10	36.74	0.32	191.6	523	MN046883
Amorphophallus titanum	4.7	18.02	36.65	0.34	192.2	470	MT161481
Arisarum simorrhinum	3.82	14.65	37.15	0.45	270.2	406	MT161482
Carlephyton glaucophyllum	4.68	17.92	37.25	0.86	513.1	738	MT161483
Steudnera colocasiifolia	3.4	13.04	37.18	0.21	131.8	234	MN046886
Typhonium blumei	5.71	21.87	34.61	2.25	1322.6	2133	MT161478
Xanthosoma helleborifolium	3.05	11.67	37.15	1.09	665.6	925	MK636779
Zomicarpella amazonica	4.54	17.42	37	0.21	129.1	508	MT161480

cp (PREP-cp) with the default setting [34]. Gene arrangement was further analyzed in the form of Collinear Block by using Mauve alignment [35] option in Geneious R8.1 [33], after removal of IRa from the genomes. The junction of cp genomes was analyzed in IRscope to visualize expansion and contraction of inverted repeats [36].

We also analyzed microsatellites repeats and oligonucleotide repeats. Microsatellites were determined by MISA [37] with the following parameters: mononucleotide ≥10, dinucleotide ≥5, trinucleotide ≥4, tetranucleotide ≥3, pentanucleotide ≥3, and hexanucleotide ≥3. We determined the forward and reverse oligonucleotide repeats by REPuter [38]. The parameters were set to search for a maximum of 500 repeats of a minimum of 19 bp with 1 mismatch.

The rates of synonymous substitutions (K_s) and non-synonymous substitutions (K_a) and their ratio K_a/K_s were determined for 75 protein-coding genes. Multiple sequence alignment was generated for all protein-coding genes using Geneious R8.1 to estimate rate of evolution of each genes in DnaSP v.6.0 [39] following Abdullah et al. [8,12] using *Amorphophallus titanum* (Becc.) as a reference genome. The data were interpreted as: purifying selection ($K_a/K_s < 1$), neutral ($K_a/K_s = 1$), and positive selection ($K_a/K_s > 1$). We also used this multiple alignment to access polymorphisms of protein-coding genes, using DnaSP v.6.0 [39].

The transition (T_s) and transversion (T_v) substitutions of the protein genes were determined by concatenating protein-coding genes. After concatenations, genes of each species were pairwise aligned by MAFFT, selecting *A. titanium* as a reference genome.

2.4. Phylogeny among species of the family Araceae

The phylogeny was inferred using 16 species from the *Dracunculus* clade using *Schismatoglottis calyptrata* (Roxb.) Zoll. & Moritzi as an outgroup (Table S1). After removal of IRa, MAFFT [40] was applied to align complete cp genomes of all species. After removal of indels, this multiple sequence alignment was used to infer the phylogeny, using IQ-tree program [41,42] with default parameters, as described in Mehmood et al. [43]. The TreeDyn was used to further enhance the visualization of phylogeny [44].

3. Results

3.1. Comparison among features of chloroplast genomes in the Dracunculus clade

Chloroplast genomes of all species showed quadripartite structure in which LSC and SSC regions are separated by IRb and IRa (Fig. 2). The size of these genomes ranged from 162,424 bp (*C. esculenta*) to 176,835 bp (*Amorphophallus titanum* Becc.). The size of LSC regions ranged from 87,141 bp (*Arisarum simorrhinum* Durieu) to 95,475 bp (*Amorphophallus titanum*). The size of each IR varied between 25,131 bp (*Zomicarpella amazonica* Bogner) and 32,708 bp (*Amorphophallus titanum*). The sizes of SSC regions varied between 14,338 bp (*Carlephyton glaucophyllum* Bogner) and 23,981 bp (*Pinellia pedatisecta*) (Table 3). These data confirmed a high level of variation among the overall sizes and also among the three main regions of the cp genomes. The average GC

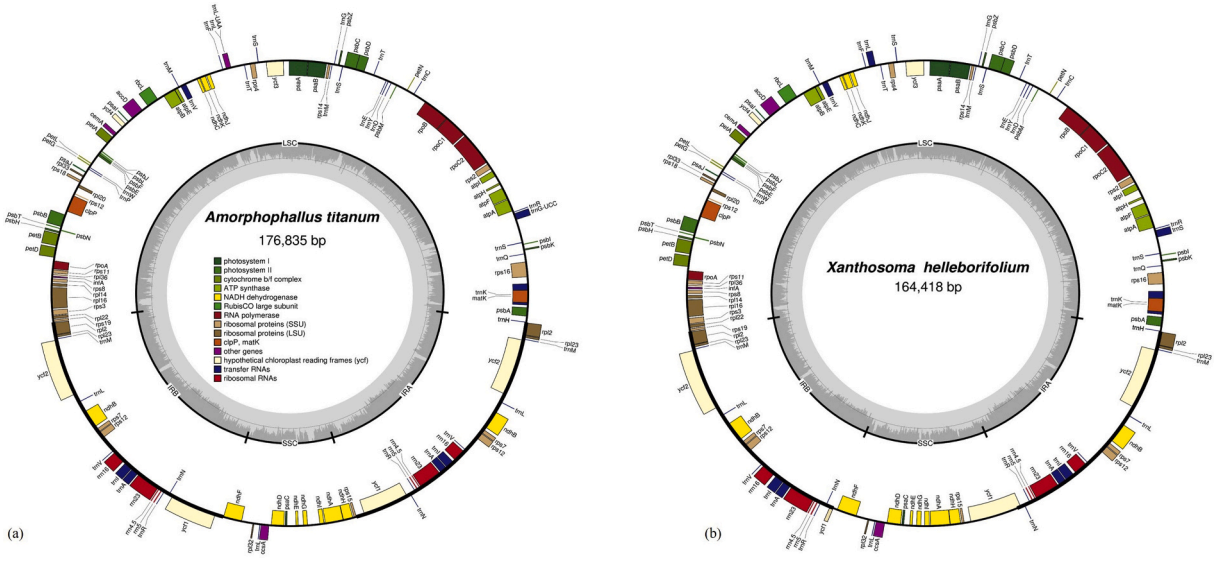

Fig. 2. Circular map of chloroplast genomes. (a) *Amorphophallus titanum* represents species that exhibit duplication of *ycf*1, whereas *Xanthosoma helleborifolium* represents species in which a single copy of *ycf*1 exists. The genes present within the circle transcribe counterclockwise and the genes present outside the circle transcribe clockwise. Genes are color-coded based on function.

Table 3
Chloroplast genomes features of 12 species of *Dracunculus* clade.

Species	Size of chloroplast genomes and their regions (bp)				GC content (%)				Genes content					
	Total	LSC	SSC	IR	Total	LSC	SSC	IR	Total	Duplicate	Protein coding	rRNA	tRNA	Pseudo copy*
Alocasia fornicata	166,052	92,316	23,213	25,262	35.7	33.8	28.7	42.4	130	17	84 (78)	8 (4)	37 (30)	*inf*A
Amorphophallus titanum	176,835	95,475	15,944	32,708	34.5	32.7	28.5	38.6	131	18	85 (78)	8 (4)	37 (30)	*inf*A
Arisarum simorrhinum	164,961	87,141	14,990	31,415	36.5	35.1	30.4	40	131	18	85 (78)	8 (4)	37 (30)	*inf*A
Carlephyton glaucophyllum	168,218	89,254	14,338	32,313	35.9	34.3	30.7	39.1	131	18	85 (78)	8 (4)	37 (30)	*inf*A
Steudnera colocasiifolia	162,500	89,806	22,196	25,249	36.1	34.4	28.9	42.5	130	17	84 (78)	8 (4)	37 (30)	*inf*A
Typhonium blumei	169,977	90,809	15,564	31,802	35.6	33.9	29.7	39.6	131	18	85 (78)	8 (4)	37 (30)	*inf*A
Xanthosoma helleborifolium	164,418	90,833	20,705	26,440	35.8	33.9	29.9	41.5	131	17	84 (78)	8 (4)	37 (30)	*inf*A,*ycf*1
Zomicarpella amazonica	162,728	90,810	21,656	25,131	35.8	34.3	28.8	41.5	131	17*	83 (78)	8 (4)	37 (30)	*inf*A, *ycf*1, *rpl*2
Arisaema franchetianum[$]	169,443	94,702	21,955	26,393	34.9	32.8	29.0	41.1	130	17	84 (78)	8 (4)	37 (30)	*inf*A
Colocasia esculenta[$]	162,424	89,670	22,208	25,273	36.2	34.4	28.9	42.4	130	17	84 (78)	8 (4)	37 (30)	*inf*A
Pinellia pedatisecta[$]	168,178	92,963	23,981	25,617	35.1	33.3	26.8	42.1	130	17	84 (78)	8 (4)	37 (30)	*inf*A
Syngonium angustatum[$]	164,929	90,714	21,559	26,328	35.7	33.9	29.0	41.5	130	17	84 (78)	8 (4)	37 (30)	*inf*A

*The *inf*A is a single copy gene and found non-functional in all species, whereas for other genes one functional copy also exist along the truncated copy. [$]The data of chloroplast genomes were retrieved from GenBank.

content in complete cp genomes was 34.5–36.5%; (Table 3).

Despite high variation in genome sizes, the analyses of gene content showed similarities among the 12 analyzed species. These genomes exhibited 113 unique genes, including 79 protein-coding genes, 30 tRNA genes, and 4 rRNA genes. The *inf*A gene was found to be non-functional in all species. Although the circular map and Mauve base collinear block analyses revealed highly similar gene arrangement in the analyzed species, variation was generated due to IR expansion and contraction (Figs. 2 and 3). The total number of genes varied among species from 130 to 131 due to IR contraction and expansion, and subsequently the generation of pseudogenes. We observed 17–18 duplicated genes in the IR, including 6–7 protein-coding genes (3 genes also included introns), 4 rRNA genes, and 7 tRNA genes (2 genes also included introns) (Table 3). Generation of pseudogenes of *ycf*1and *rpl*2 was also observed (Table 3). We found 18 intron-containing genes, of which 12 were protein-coding genes and 6 were tRNA genes. Two introns were observed in *ycf3* and *clpP,* whereas all other genes contained one intron.

3.2. Contraction and expansion of inverted repeats

The contraction and expansion phenomena of IRs led to a variable number of total genes. This variation was due to conversion of duplicate genes to single copy or *vice versa*. Moreover, contraction and expansion of IRs also produced pseudogenes. The junctions JLB (LSC/IRb) and JLA (IRa/SSC) showed high similarities and the existence of *rps*19 & *rpl*2 and *rpl*2 & *trnH*-GUG, respectively. In *Z. amazonica*, the IR contraction at IRs/LSC resulted in a pseudo-copy of *rpl*2 at the junction of JLB; whereas at JLA the complete intron and exon 2 of *rpl*2 were transferred to the LSC region and the *trn*H gene was present 1148 bp away from the junction of JLA. In other species, *trn*H was found to be at JLA, either starting inside the IRa (6 to 21 bases) or starting up to 20 bases after the start of LSC region. The junctions JSB (IRb/SSC) and JSA (SSC/IRa) were highly variable. We noticed the complete duplication of *ycf*1 in *Amorphophallus titanum*, *Arisarum simorrhinum*, *C. glaucophyllum*, and *Typhonium blumei* Nicolson & Sivad. due to IR expansion (Fig. 4). Consequently, *ycf*1 and *ndh*F were present at JSB, and *rps*15 and *ycf*1 were present at JSA in these four species. A pseudo-copy of *ycf*1 gene was found at the junction of JSB along with a functional copy of *ycf*1 in *Xanthosoma helleborifolium* (Jacq.) Schott, *S. angustatum*, and *Z. amazonica* (Fig. 4). As a result, *ycf*1 pseudo-copy and *ndh*F genes exist at JSB, whereas the functional copy of *ycf*1 starts from IR and enters SSC regions at JSA in these species. In another five species: *P. pedatisecta*, *C. esculenta*, *A. franchetianum*, *Alocasia fornicata* (Roxb.) Schott, and *Steudnera colocasiifolia* K.Koch, the *ycf*1 gene completely exists in SSC regions (Fig. 4). Hence, *trn*N and *ndh*F genes were found at JSB (Fig. 4). The *ndh*F gene showed integration at the junction from SSC to IRb in *A. franchetianum*, *T. blumei*, and *C. glaucophyllum*. A functional copy and a pseudo-copy of *rps*15 gene were found in in *C. glaucophyllum* due to IR expansion (Fig. 4).

3.3. Analyses of codon usage, amino acid frequency, and RNA editing sites

The relative synonymous codon usage analyses showed high similarities among species and a trend towards A/T content. We recorded high codon usage of codons that contain A/T at 3′ as compared with codons that contain C/G at 3′ (Table S2). We recorded high similarity in the amino acid frequency among the species; leucine and isoleucine were the most frequently coded amino acids and cysteine was the least coded amino acid (Table S3). The RNA editing analyses also showed similarities with respect to genes and position of editing sites in the coding genes. We recoded RNA editing sites in 18 to 22 genes among different species (Table S4). The *rpl*2 genes contained ACG as a start codon instead of ATG. The RNA editing analyses confirmed conversion of ACG to ATG.

3.4. Repeats analyses

Variable number of microsatellites repeats (80 to 209) were observed in the analyzed species. We recorded all six types of microsatellites, in which most repeats were composed of A/T motifs (Table S5). Most repeats were located in LSC regions, followed by SSC and IR (Fig. S1a). Mononucleotide repeats were abundant relative to other types of repeats

Fig. 3. Collinear block analyses based on Mauve alignment show rearrangement of genes existing at the junction of inverted repeat and small-single copy. White block: protein-coding genes, black block: tRNA genes, green block: intron containing tRNA genes, red block: rRNA genes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. S1b). Analysis of oligonucleotide repeats showed high fluctuations in repeats numbers among the species, ranging from 208 to 491. The forward repeats were more abundant than the reverse repeats in eleven species, while reverse repeats exceeded the number of forward repeats in *T. blumei* and *Z. amazonica.* (Fig. S2a). The number of forward repeats varied from 119 to 296, whereas the number of reverse repeats varied from 89 to 236. Repeats mostly existed in LSC regions, whereas SSC and IR showed fluctuations and were not consistent regarding repeats numbers (Fig. S2b). Higher numbers of repeats were found in those IR in which part of the SSC was transferred to IR. We also observed some shared repeats in LSC/IR, LSC/SSC, and SSC/IR, in which one copy of a repeat was found in one region and another copy of a repeat in other regions (Fig. S2b). Most repeats ranged from 20 bp to 24 bp (Fig. S2c). Details about their size, types, and positions in the cp genome have been provided in Table S6.

3.5. Rates of synonymous and non-synonymous substitutions and transition and transversion substitutions

With a few exceptional genes mentioned below, the ratio of non-synonymous substitutions (K_a) to synonymous substitutions (K_s) revealed purifying selection for protein-coding genes ($K_a/K_s < 1$; Table S7). The genes of photosystem I and photosystem II were more conserved and showed higher purifying selection when compared to other genes. Positive selection was observed for *rpl*33 (*T. blumei*), *rps*8 (*X. helleborifolium*), *rps*16 (*Z. amazonica*), and *ycf*2 genes (*A. fornicata*, *C. esculenta*, *S. colocasiifolia*, *A. franchetianum*, *A. simorrhinum*, and *C. glaucophyllum*). Neutral selection was observed in *clp*P, *rpl*16, *rps*8, and *ycf*2 genes in some species (Table S7). We also analyzed transition (Ts) and transversion (T_v) mutations in the protein-coding genes. Transitions were higher than transversions; T_s/T_v ratio ranged from 2.29 to 2.89 (Table 4).

3.6. Identification of suitable polymorphic loci for phylogenetic inference

For phylogenetic inference of distantly related (deeply diverged) species, relatively slow evolving coding regions having adequate levels of polymorphism and should be preferred over mutational hotspots. We identified 10 suitable polymorphic coding sequences for phylogenetic inference (Table 5). The nucleotide diversity of these loci ranged from 0.0315 to 0.0235. These loci include *ndh*G, *mat*K, *rps*19, *ndh*F, *rps*11, *ccs*A, and *rps*8 genes.

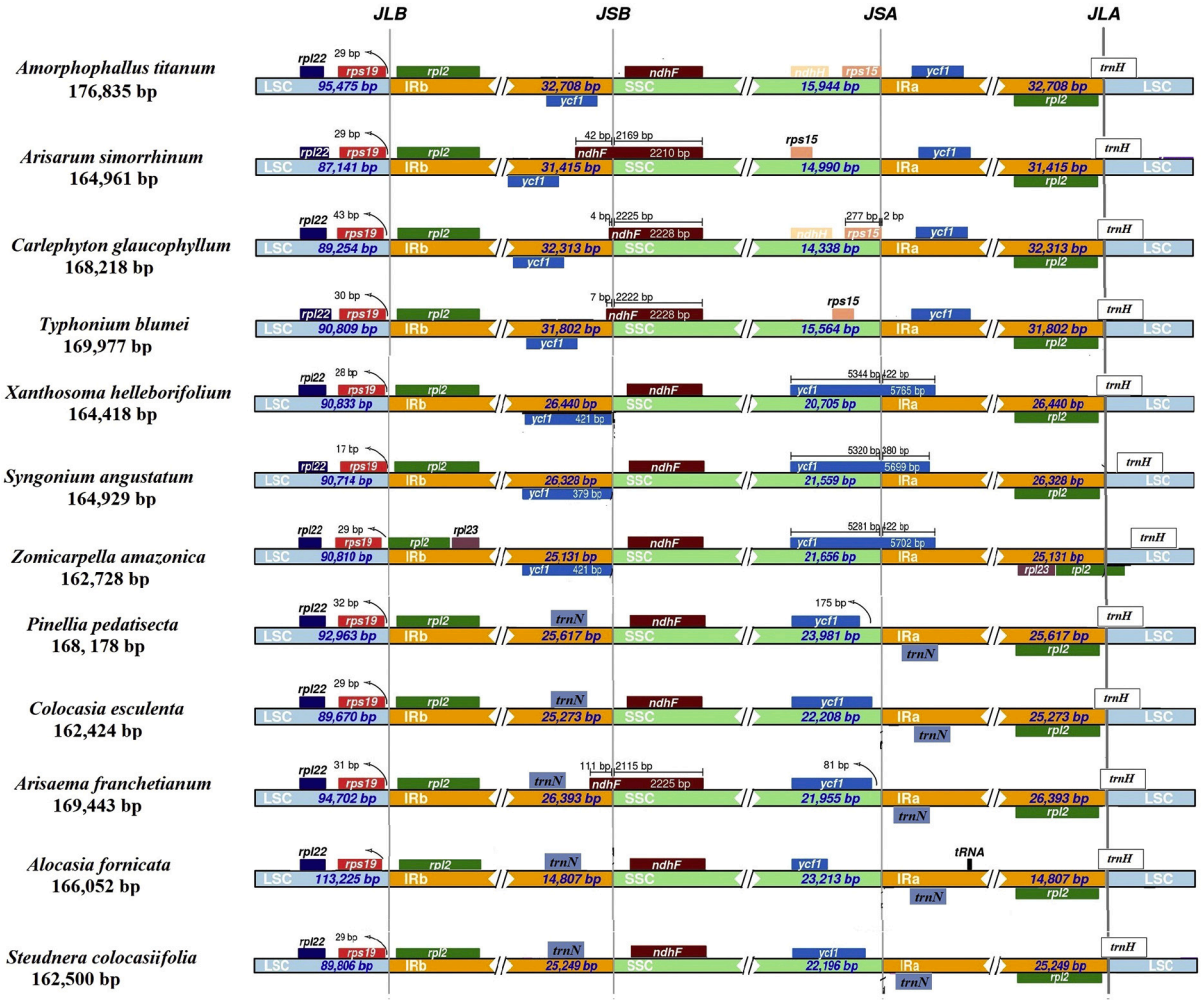

Fig. 4. Analyses of expansion and contraction of inverted repeats. The genes present at the top of track transcribe counterclockwise, whereas genes present below the track transcribe clockwise.

Table 4
Comparison of transition and transversion substitutions among 12 species of *Dracunculus* clade.

Species	C/T	A/G	A/T	A/C	G/T	C/G	T_s/T_v
Alocasia fornicata	524	518	81	169	115	37	2.59
Arisarum simorrhinum	553	547	82	141	114	44	2.89
Carlephyton glaucophyllum	476	451	93	134	92	28	2.67
Steudnera colocasiifolia	497	486	81	156	120	36	2.50
Typhonium blumei	598	599	107	180	128	37	2.65
Xanthosoma helleborifolium	406	415	80	148	104	27	2.29
Zomicarpella amazonica	392	384	70	139	97	28	2.32
Arisaema franchetianum	580	603	100	198	156	39	2.40
Colocasia esculenta	489	490	86	157	117	38	2.46
Pinellia pedatisecta	550	567	101	169	129	35	2.57
Syngonium angustatum	385	391	58	131	94	26	2.51

Table 5
Suitable polymorphic loci for phylogenetic inference.

S. No	Region	Nucleotide Diversity	Total number of mutations	Region Length
1	*matK*	0.0315	202	1563
2	*rps*19	0.0303	28	288
3	*ndh*G	0.0288	51	531
4	*ndh*F	0.0285	250	2249
5	*rps*11	0.0277	48	417
6	*ccs*A	0.0267	106	972
7	*rps*8	0.0267	38	399
8	*acc*D	0.0257	152	1614
9	*ndh*D	0.0244	146	1530
10	*rpl*33	0.0235	23	201

3.7. Phylogeny of the family Araceae

The phylogeny was inferred among 16 species of *Dracunculus* clade using *Schismatoglottis calyptrata* (Roxb.) Zoll. & Moritzi as outgroup with the best fit model K3Pu + F + I + G4. The multiple alignment which we used for phylogenetic inference contained 112,119 total nucleotide sites after removal of InDels. The phylogeny among the species supports the monophyly of *Dracunculus* clade as well as the subclades *Amorphophallus*, *Colletogyne*, and *Pistia*. A close relationship was observed between *S. colocasiifolia* and *C. esculenta* (Fig. 5). The relationships among the species was supported within clade with high bootstrapping value of 100. However, the bootstrapping value remained lower (70) for the node separating *Typhonium* from *Pinellia* and *Arisaema*. Another short internal branch, and hence low bootstrap score (84) was found for the node splitting *Colocasia/Steudnera* from *Alocasia* (Fig. 5).

4. Discussion

We report *de novo* assembled cp genomes of eight species from the *Dracunculus* clade (Aroideae, Araceae). These species show variation in cp genome size and total numbers of genes. The number of total genes

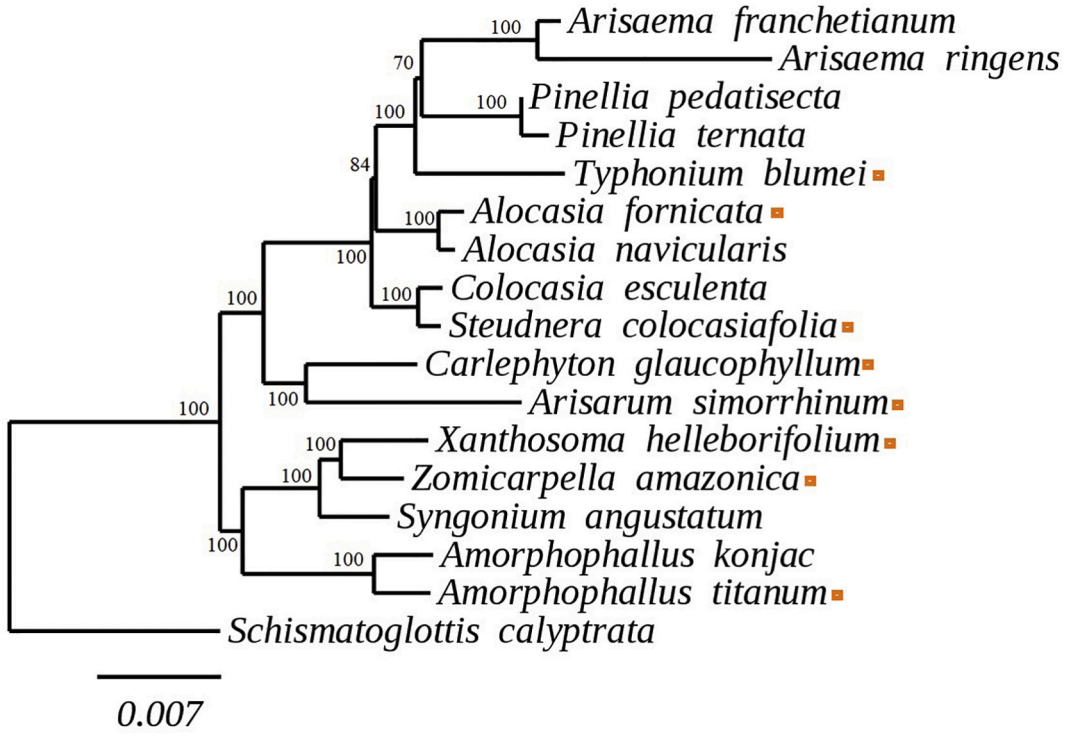

Fig. 5. Phylogenetic inference among aroid species based on multiple alignment of complete chloroplast genome by IQ-tree. The chloroplast genomes of marked species were assembled in current study.

varied due to IR contraction and expansion, which led to duplication as well as truncation of some genes, but retained at least one functional copy of all genes. In terms of quadripartite genome structure, we observed highly conserved cp genomes in the 12 analyzed species. These results are in agreement with previous reports of the species of seven subfamilies of Araceae [19,20,25,45–47]. Moreover, *A. titanum* was found to be the largest cp genome among the reported species.

Previously, a loss of some important genes, including deletion of *ycf*1 in *Amorphophallus bulbifer* (Roxb.) Blume, *trnG-GCC* in *Amorphophallus konjac* K.Koch, *trn*L-*CCA* in *Amorphophallus bulbifer* and *Amorphophallus muelleri* Blume, *psb*E in *Amorphophallus konjac*, and *accD* in *Amorphophallus albus* P.Y.Liu & J.F.Chen, *Amorphophallus konjac,* and *Amorphophallus bulbifer* was reported by Liu et al. [21]. They also observed deletion of *rpl*2 and *rpl*23 from IRb, whereas functional copies existed in IRa. In our study, a copy of *rpl*2 was truncated at the JLB (LSC/IRb) junction due to contraction in IR region in *Zomicarpella amazonica*, but not in *Amorphophallus titanum*. Keeping in view the highly conservative nature of cp genes across the *Dracunculus* clade, we propose that the deletions reported by Liu et al. [21] can be attributed to errors in either the assembly, or the annotations, or both. Liu et al. [21] used Dual Organellar Genome Annotator (DOGMA) for annotations, which is error prone [48]. Our critique is consistent with the recent report of the cp genome of *Amorphophallus konjac* (MK611803.1) [49], which supports the existence of all of the genes in the cp genome. In contrast, Liu et al. [21] reported deletion of *accD, psbE, and trnG-GCC* in their cp genome of *A. konjac*. The cp genome of *A. titanium*, which we sequenced and assembled *de novo* in this study, shows similar gene content to previous reports in aroids as well as to *A. konjac* (MK611803.1).

Previous studies have shown that IR contraction and expansion can lead to variable numbers of genes by converting duplicate genes to single copy and *vice versa* in aroids as well as in other angiosperms [4,19,20,32]. In the current study, *A. titanum*, *A. simorrhinum*, *C. glaucophyllum*, and *T. blumei* of the *Dracunculus* clade showed two functional copies of *ycf*1 gene due to duplication in IR regions, while the rest of the species contained one functional copy and a pseudo-gene (truncated copy) of this gene in IRb region. We previously reported transfer of *ycf*1 gene to SSC region, duplication with two functional copies in IR regions, or one functional copy and a truncated copy (pseudogene) in different subfamilies in the family Araceae. For example, in subfamily Aroideae, *ycf*1 gene was completely transferred to SSC region, and hence exists as single copy with no truncated copy in *Colocasia* [32], *Alocasia*, *Calla, Arisaema, Schismatoglottis, Pinellia, Taccarum* and *Montrichardia* species in subfamily Aroideae [19]. In *Anubias, Aglaonema, Syngonium and Philodendron* species in subfamily Aroideae, the functional copy of this gene extends into IRa from SSC, hence a truncated copy also exists in IRb [19]. In *Zantedeschia* and *Anchomanes* species in the Aroideae subfamily, two functional copies of this gene exist, one each in IRa and IRb regions [19]. In subfamily Pothoideae [50], Orontioideae, Lasioideae, Zamioculcadoideae [20], and Lemnoideae [45], this gene exists as two functional copies, one each in IRa and IRb. In Monsteroideae subfamily, this gene either exists fully in SSC or extended in IRb [25]. This suggests that the duplication events of *ycf*1 and/or other genes that exist on junctions of the single-copy and inverted repeats in cp genome are species-specific and not synapomorphies for clades. Further genomic resources may provide better insight into the phylogenetic levels at which IR contraction and expansion may be of diagnostic utility in Araceae.

In this study, we report polymorphic loci among the coding sequences of *ndh*G, *mat*K, *rps*19, *ndh*F, *rps*11, *ccs*A, and *rps*8 that could potentially provide high resolution for phylogenetic studies among the genera within the *Dracunculus* clade, specifically for those with taxonomic issues [1,51,52]. We ignored the intergenic spacer regions, as such regions showed rapid evolution and could be misleading in drawing phylogenetic inferences. We also ignored *ycf*1 and *rps*15 due to effect of rate heterotachy as reported previously [50]. Our phylogenetic inference also agreed with previous studies of Araceae and showed similar relationships among the species of *Dracunculus* clade [15,18,53].

In conclusion, our study supports the existence of a conserved cp genome structure in the *Dracunculus* clade and the family Araceae, with a variable number of total genes due to contraction and expansion of IRs.

The identified polymorphic loci will facilitate phylogenetic inference of the *Dracunculus* clade and its genera.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.12.016.

Author contributions

Sample collection, DNA extraction, and sequencing: C.L.H. and T. B·C.

Genome assembly: A, S·W, and C.L.H.

Coverage depth analyses and annotations: A and C.L.H.

Data analyses: A, F.M., A.S, and A.H.

Data interpretation: A, P·P, and MTW.

Conceptualization: A., P·P., I.A., P.J.M and T.B·C.

Visualization: A, P.P, P.J.M.

Data curation: A., FM, and C.L.H.

project administration: A. and C.L.H.

Writing—original draft: A.

Writing—review and editing: C.L.H., I.A., P.P., P.JM,

Supervision: P.P. and I.A.

All authors have read and agreed to the published version of the manuscript.

Funding

Funding for this study was provided by the by the Graduate Assistance in Areas of National Need (GAANN) fellowship, the Rettner B. Morris Scholarship, Washington University in St. Louis, and the J. Chris Pires Lab (NSF DEB 1146603).

Declaration of Competing Interest

No conflicts of interest exist.

Acknowledgments

The authors are grateful for funding and laboratory support from Barbara Schaal at Washington University in St. Louis and J. Chris Pires at the University of Columbia, Missouri. The authors thank Tatiana Arias for valuable help in the laboratory and data processing. In the aroid greenhouse at the Missouri Botanical Garden, Emily Colletti provided critical assistance with living material.

References

[1] H. Daniell, C.-S. Lin, M. Yu, W.-J. Chang, Chloroplast genomes: diversity, evolution, and applications in genetic engineering, Genome Biol. 17 (2016) 134, https://doi.org/10.1186/s13059-016-1004-2.

[2] J.D. Palmer, Comparative organization of chloroplast genomes, Annu. Rev. Genet. 19 (1985) 325–354, https://doi.org/10.1146/annurev.ge.19.120185.001545.

[3] S. Iram, M.Q. Hayat, M. Tahir, A. Gul, Abdullah, I. Ahmed, Chloroplast genome sequence of *Artemisia scoparia*: comparative analyses and screening of mutational hotspots, Plants. 8 (2019) 476, https://doi.org/10.3390/plants8110476.

[4] I. Shahzadi, Abdullah, F. Mehmood, Z. Ali, I. Ahmed, B. Mirza, Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus *Artemisia* and phylogeny in family Asteraceae, Genomics. 112 (2020) 1454–1463, https://doi.org/10.1016/J.YGENO.2019.08.016.

[5] Y. Zhang, J. Ma, B. Yang, R. Li, W. Zhu, L. Sun, J. Tian, L. Zhang, The complete chloroplast genome sequence of *Taxus chinensis* var. *mairei* (Taxaceae): loss of an inverted repeat region and comparative analysis with related species, Gene. 540 (2014) 201–209, https://doi.org/10.1016/j.gene.2014.02.037.

[6] U. Zeb, W. Dong, T. Zhang, R. Wang, K. Shahzad, X. Ma, Z. Li, Comparative plastid genomics of *Pinus* species: Insights into sequence variations and phylogenetic relationships, J. Syst. Evol. (2019), https://doi.org/10.1111/jse.12492 jse.12492.

[7] D.J. Oldenburg, A.J. Bendich, The linear plastid chromosomes of maize: terminal sequences, structures, and implications for DNA replication, Curr. Genet. 62 (2016) 431–442, https://doi.org/10.1007/s00294-015-0548-0.

[8] Abdullah, I. Shahzadi, F. Mehmood, Z. Ali, M.S Malik, S. Waseem, B. Mirza, I. Ahmed, M.T Waheed, Comparative analyses of chloroplast genomes among three *Firmiana* species: identification of mutational hotspots and phylogenetic relationship with other species of Malvaceae, Plant Gene. 19 (2019), 100199, https://doi.org/10.1016/J.PLGENE.2019.100199.

[9] F. Mehmood, Abdullah, Z. Ubaid, Y. Bao, P. Poczai, B. Mirza, Comparative plastomics of Ashwagandha (Withania, Solanaceae) and identification of mutational hotspots for barcoding medicinal plants, Plants 9 (2020) 752, https://doi.org/10.20944/preprints202003.0181.v1.

[10] F. Mehmood, Abdullah, I. Shahzadi, I Ahmed, M.T Waheed, B. Mirza, Characterization of *Withania somnifera* chloroplast genome and its comparison with other selected species of Solanaceae, Genomics. 112 (2020) 1522–1530, https://doi.org/10.1016/j.ygeno.2019.08.024.

[11] Abdullah, F. Mehmood, I. Shahzadi, Z. Ali, M. Islam, M. Naeem, B. Mirza, P. Lockhart, I. Ahmed, M.T. Waheed, Correlations among oligonucleotide repeats, nucleotide substitutions and insertion-deletion mutations in chloroplast genomes of plant family Malvaceae, J. Syst. Evol. (2020), https://doi.org/10.1111/jse.12585.

[12] Abdullah, F. Mehmood, S. Shahzadi, S. Waseem, B. Mirza, I. Ahmed, M.T. Waheed, Chloroplast genome of *Hibiscus rosa-sinensis* (Malvaceae): comparative analyses and identification of mutational hotspots, Genomics. 112 (2020) 581–591, https://doi.org/10.1016/j.ygeno.2019.04.010.

[13] I. Ahmed, P.J. Matthews, P.J. Biggs, M. Naeem, P.A. Mclenachan, P.J. Lockhart, Identification of chloroplast genome loci suitable for high-resolution phylogeographic studies of *Colocasia esculenta* (L.) Schott (Araceae) and closely related taxa, Mol. Ecol. Resour. 13 (2013) 929–937, https://doi.org/10.1111/1755-0998.12128.

[14] P.C. Boyce, T.B. Croat, The Überlist of Araceae, Totals for Published and Estimated Number of Species in Aroid Genera, 2018.

[15] N. Cusimano, J. Bogner, S.J. Mayo, P.C. Boyce, S.Y. Wong, M. Hesse, W.L. A. Hetterscheid, R.C. Keating, J.C. French, Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies, Am. J. Bot. 98 (2011) 654–668, https://doi.org/10.3732/ajb.1000158.

[16] M.H. Grayum, Evolution and phylogeny of the Araceae, Ann. Missouri Bot. Gard. 77 (1990) 628, https://doi.org/10.2307/2399668.

[17] S.J. Mayo, J. Bogner, E. Catherine, P.J. Boyce, The Genera of Araceae, Royal Botanic Gardens, Kew, [London], 1997.

[18] C.L. Henriquez, T. Arias, J.C. Pires, T.B. Croat, B.A. Schaal, Phylogenomics of the plant family Araceae, Mol. Phylogenet. Evol. 75 (2014) 91–102, https://doi.org/10.1016/j.ympev.2014.02.017.

[19] C.L. Henriquez, Abdullah, I. Ahmed, M.M. Carlsen, A. Zuluaga, T.B. Croat, M. R. Mckain, Evolutionary dynamics of chloroplast genomes in subfamily Aroideae (Araceae), Genomics. 112 (2020) 2349–2360, https://doi.org/10.1016/j.ygeno.2020.01.006.

[20] Abdullah, C.L. Henriquez, F. Mehmood, I. Shahzadi, Z. Ali, M.T. Waheed, T. B. Croat, P. Poczai, I. Ahmed, Comparison of chloroplast genomes among species of unisexual and bisexual clades of the monocot family Araceae, Plants. 9 (2020) 737, https://doi.org/10.3390/plants9060737.

[21] E. Liu, C. Yang, J. Liu, S. Jin, N. Harijati, Z. Hu, Y. Diao, L. Zhao, Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species, Sci. Rep. 9 (2019) 809, https://doi.org/10.1038/s41598-018-37456-z.

[22] R.C. Keating, Vegetative anatomical data and its relationship to a revised classification of the genera of Araceae, Ann. Missouri Bot. Gard. 91 (2004) 485–494, https://doi.org/10.2307/3298625.

[23] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2017. (Accessed 15 September 2019).

[24] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics. 32 (2016) 3047–3048, https://doi.org/10.1093/bioinformatics/btw354.

[25] C.L. Henriquez, Abdullah, I. Ahmed, M.M. Carlsen, A. Zuluaga, T.B. Croat, M. R. Mckain, Molecular evolution of chloroplast genomes in Monsteroideae (Araceae), Planta. 251 (2020) 72, https://doi.org/10.1007/s00425-020-03365-7.

[26] B. Langmead, C. Trapnell, M. Pop, S. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25, https://doi.org/10.1186/gb-2009-10-3-r25.

[27] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq – versatile and accurate annotation of organelle genomes, Nucleic Acids Res. 45 (2017) W6–W11, https://doi.org/10.1093/nar/gkx391.

[28] T.M. Lowe, P.P. Chan, tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes, Nucleic Acids Res. 44 (2016) W54–W57, https://doi.org/10.1093/nar/gkw413.

[29] D. Laslett, B. Canback, ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences, Nucleic Acids Res. 32 (2004) 11–16, https://doi.org/10.1093/nar/gkh152.

[30] S. Greiner, P. Lehwark, R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes, Nucleic Acids Res. 47 (2019) W59–W64, https://doi.org/10.1093/nar/gkz238.

[31] P. Lehwark, S. Greiner, GB2sequin - a file converter preparing custom GenBank files for database submission, Genomics. 111 (2019) 759–761, https://doi.org/10.1016/J.YGENO.2018.05.003.

[32] I. Ahmed, P.J. Biggs, P.J. Matthews, L.J. Collins, M.D. Hendy, P.J. Lockhart, Mutational dynamics of aroid chloroplast genomes, Genome Biol. Evol. 4 (2012) 1316–1323, https://doi.org/10.1093/gbe/evs110.

[33] M. Kearse, R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond, Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics. 28 (2012) 1647–1649, https://doi.org/10.1093/bioinformatics/bts199.

[34] J.P. Mower, The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments, Nucleic Acids Res. 37 (2009) W253–W259, https://doi.org/10.1093/nar/gkp337.
[35] A.C.E. Darling, B. Mau, F.R. Blattner, N.T. Perna, Mauve: multiple alignment of conserved genomic sequence with rearrangements, Genome Res. 14 (2004) 1394–1403, https://doi.org/10.1101/gr.2289704.
[36] A. Amiryousefi, J. Hyvönen, P. Poczai, IRscope: an online program to visualize the junction sites of chloroplast genomes, Bioinformatics. 34 (2018) 3030–3031, https://doi.org/10.1093/bioinformatics/bty220.
[37] S. Beier, T. Thiel, T. Münch, U. Scholz, M. Mascher, MISA-web: a web server for microsatellite prediction, Bioinformatics. 33 (2017) 2583–2585, https://doi.org/10.1093/bioinformatics/btx198.
[38] S. Kurtz, J.V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, R. Giegerich, REPuter: the manifold applications of repeat analysis on a genomic scale, Nucleic Acids Res. 29 (2001) 4633–4642.
[39] J. Rozas, A. Ferrer-Mata, J.C. Sanchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E. Ramos-Onsins, A. Sanchez-Gracia, DnaSP 6: DNA sequence polymorphism analysis of large data sets, Mol. Biol. Evol. 34 (2017) 3299–3302, https://doi.org/10.1093/molbev/msx248.
[40] K. Katoh, K.I. Kuma, H. Toh, T. Miyata, MAFFT version 5: improvement in accuracy of multiple sequence alignment, Nucleic Acids Res. 33 (2005) 511–518, https://doi.org/10.1093/nar/gki198.
[41] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274, https://doi.org/10.1093/molbev/msu300.
[42] D.T. Hoang, O. Chernomor, A. von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2018) 518–522, https://doi.org/10.1093/molbev/msx281.
[43] F. Mehmood, Abdullah, Z. Ubaid, I. Shahzadi, I. Ahmed, M.T. Waheed, P. Poczai, B. Mirza, Plastid genomics of *Nicotiana* (Solanaceae): insights into molecular evolution, positive selection and the origin of the maternal genome of Aztec tobacco (*Nicotiana rustica*), BioRxiv (2020), https://doi.org/10.1101/2020.01.13.905158.
[44] F. Chevenet, C. Brun, A.-L. Bañuls, B. Jacq, R. Christen, TreeDyn: towards dynamic graphics and annotations for analyses of trees, BMC Bioinformeatics. 7 (2006) 439, https://doi.org/10.1186/1471-2105-7-439.
[45] W. Wang, J. Messing, High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA, PLoS One 6 (2011), https://doi.org/10.1371/journal.pone.0024670.
[46] S.-H. Kim, J. Yang, J. Park, T. Yamada, M. Maki, S.-C. Kim, Comparison of whole plastome sequences between Thermogenic skunk cabbage *Symplocarpus renifolius* and Nonthermogenic *S. nipponicus* (Orontioideae; Araceae) in East Asia, Int. J. Mol. Sci. 20 (2019) 4678, https://doi.org/10.3390/ijms20194678.
[47] K.S. Choi, K.T. Park, S. Park, The chloroplast genome of *Symplocarpus renifolius*: comparison of chloroplast genome structure in Araceae, Genes (Basel) 8 (2017) 324, https://doi.org/10.3390/genes8110324.
[48] A. Amiryousefi, J. Hyvönen, P. Poczai, The chloroplast genome sequence of bittersweet (*Solanum dulcamara*): plastid genome structure evolution in Solanaceae, PLoS One 13 (2018) 1–23, https://doi.org/10.1371/journal.pone.0196069.
[49] H. Hu, J. Liu, B. Wang, J. An, Q. Wang, Characterization of the complete chloroplast genome of *Amorphophallus konjac* (Araceae) and its phylogenetic analysis, Mitochondrial DNA Part B Resour. 4 (2019) 1658–1659, https://doi.org/10.1080/23802359.2019.1606683.
[50] Abdullah, C.L. Henriquez, F. Mehmood, M.M. Carlsen, M. Islam, M.T. Waheed, P. Poczai, T.B. Croat, I. Ahmed, Ahmed, complete chloroplast genomes of *Anthurium huixtlense* and *Pothos scandens* (Pothoideae, Araceae): unique inverted repeat expansion and contraction affect rate of evolution, BioRxiv. (2020), https://doi.org/10.1101/2020.03.11.987859.
[51] I. Ahmed, Evolutionary Dynamics in Taro, PhD Thesis, Massey University, Palmerston North, New Zealand, 2014, https://mro.massey.ac.nz/handle/10179/5610.
[52] Abdullah, S. Waseem, B. Mirza, I. Ahmed, M.T. Waheed, Comparative analyses of chloroplast genomes of *Theobroma cacao* and *Theobroma grandiflorum*, Biologia (Bratisl). 75 (2020) 761–771, https://doi.org/10.2478/s11756-019-00388-8.
[53] L. Nauheimer, D. Metzler, S.S. Renner, Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils, New Phytol. 195 (2012) 938–950, https://doi.org/10.1111/j.1469-8137.2012.04220.x.